

基于核密度互信息的无监督特征选择方法

袁 松

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2023年10月23日; 录用日期: 2023年12月20日; 发布日期: 2023年12月28日

摘 要

特征选择是对高维数据进行分析的基础, 是许多模式识别应用程序中的一个关键过程。由于使用大量的特征, 学习模型会出现过拟合, 导致模型性能下降等, 因此对高维数据进行特征筛选变得尤为重要。本文提出基于核密度互信息的无监督特征选择方法, 算法引入互信息作为特征之间的相关性度量, 可以很好地度量特征间的线性 and 非线性关系, 同时采用核密度估计的方式估计特征的密度函数, 可以对互信息更准确地计算。最后将该方法应用在两个真实的高维数据集上, 并将筛选后的数据应用于三个常用的分类模型, 采用分类精度和F1值对分类模型进行评价。通过与现有的基于互相关性方法作对比分析得知, 在两个数据集上, 两种筛选方法在三个分类模型中的分类精度表现类似, 而基于核密度互信息的方法在三种分类算法上的F1值比基于互相关性的方法表现更好, 表明本文所提方法具有更好的筛选性能。

关键词

高维数据, 特征选择, 无监督, 互信息, 核密度

Unsupervised Feature Selection Method Based on Kernel Density Mutual Information

Song Yuan

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Oct. 23rd, 2023; accepted: Dec. 20th, 2023; published: Dec. 28th, 2023

Abstract

Feature selection is fundamental to the analysis of high-dimensional data and is a key process in many pattern recognition applications. Due to the use of a large number of features, the learning model will overfit, resulting in the degradation of model performance, etc., so the feature screening of high-dimensional data becomes particularly important. This paper proposes an unsupervised

feature selection method based on kernel density mutual information. Mutual information is introduced as the correlation measure between features, which can well measure the linear and nonlinear relationship between features. Finally, the method is applied to two real high-dimensional data sets, and the filtered data is applied to three commonly used classification models, and the classification model is evaluated by classification accuracy and F1 value. Through comparative analysis with the existing cross-correlation-based method, it is known that on the two data sets, the classification accuracy performance of the two screening methods in the three classification models is similar, and the F1 value of the method based on kernel density mutual information is better than the method based on cross-correlation on the three classification algorithms, indicating that the proposed method has better screening performance.

Keywords

High-Dimensional Data, Feature Selection, Unsupervised, Mutual Information, Kernel Density

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着大数据时代的到来,以及计算能力和其他现代技术的快速发展,在当代许多研究中经常出现复杂的高维数据。在遗传学、生物学等各种科学研究领域存在大量高维数据。对于传统的学习模型,由于使用大量的特征,模型往往会过拟合,导致模型性能下降。而且过多的数据也会显著增加数据分析的内存需求和计算成本。降维是解决前面描述的问题的最强大的工具之一,它主要可分为特征提取和特征选择两个主要组成部分[1]。特征提取和特征选择都具有提高学习性能、提高计算效率、减少内存存储、建立更好的泛化模型等优点,因此都被认为是有效的降维技术。相对于特征提取,特征选择保持了原始特征的物理意义,并给模型提供了更好的可读性和可解释性,因此,特征选择在各种数据挖掘和分析中有更多的应用。通过特征选择去除这些特征可以减少存储和计算成本,同时避免了信息的重大损失或学习性能的下降。

特征选择方法根据数据集中可用的信息,可分为监督、半监督和无监督[2]的方法。监督的方法需要标记数据来识别和选择相关特征,半监督方法只要求一些数据被标记,而无监督特征选择方法则不需要被标记的数据集。对于无监督的特征选择方法,常用的有三种方法,即包装方法、嵌入式方法和过滤方法[3]。包装的特征选择方法的选择取决于分类器的性能,该方法直接利用分类算法对特征选择得到的特征子集进行评价,目的是选择最有利于给定分类器性能的特征子集。对于包装的方法,Hancer [4]等人提出了一种基于变串长度的多目标差分进化方法(MODE-CFS)来同时进行聚类 and 特征选择,为了评估解的优劣,MODE-CFS 在最小化 WB 指数的同时最大化 Silhouette 指数和所选特征的数量。嵌入式的特征选择方法也取决于分类器,它将特征选择过程与学习者训练过程相结合,两者在同一优化过程中完成,即在分类器训练过程中自动进行特征选择。对于嵌入式的方法,Zhu [5]等人提出了判别嵌入无监督特征选择方法(DEUFS),该方法通过最大化聚类之间的异质性来获取聚类标签,从而对数据的聚类结构进行建模。过滤的特征选择方法独立于特定的分类器,该方法首先选择数据集的特征,然后训练分类器。对于过滤的特征选择方法,Solorio-Fernández [6]等人介绍了一种用于混合数据的单变量无监督光谱特征选择方法,称为 USFSM(混合数据的无监督光谱特征选择方法)。USFSM 通过分析当每个特征分别从整个特

征集中排除时, 归一化拉普拉斯矩阵的第一个非平凡特征值的谱分布(谱间隙)的变化来评估特征, 最后特征根据各自的光谱间隙值按降序排序。由于包装器特征选择方法是针对给定的分类器进行直接优化的, 因此包装器特征选择方法的性能优于过滤的方法, 但包装器的计算成本通常远高于过滤的方法, 且过滤的方法很容易扩展到高维数据集, 因此具有计算简单、快速, 计算成本低和可拓展性等优点。

Haindl [7]等人提出了一种基于过滤的特征选择方法, 引入了一种基于互相关性(Mutual Correlation)的特征选择方法。该方法首先采用皮尔逊相关系数计算数据中特征之间所有可能的成对相关性, 然后删除与其他特征平均依赖性最高的特征, 对于所有特性, 重复此过程, 直到选择需要的特征数量, 他们的实验结果表明, 虽然筛选效果不如包装的方法, 但也实现了较好的筛选, 且训练更快。然而皮尔逊相关系数只能度量变量间的线性相关性, 会忽略变量间的非线性相关性。互信息是对一个随机变量包含另一个随机变量的信息量的度量, 也可以描述为在了解另一个随机变量的情况下, 降低了原始随机变量的不确定性, 互信息可以很好的度量特征间的线性和非线性关系。本文利用互信息的概念来度量特征和特征之间的关系, 建立基于互信息的无监督特征选择方法。

2. 特征选择方法

2.1. 互信息的估计

互信息[8]是两个随机变量之间的相互依赖性度量, 对于任意两个连续随机变量 X 和 Y , 其互信息为:

$$I(X, Y) = \iint p(x, y) \log_2 \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy \quad (1)$$

其中 $p(x), p(y)$ 分别为 x, y 的概率密度函数, $p(x, y)$ 是 x 和 y 的联合概率密度函数。对于任意两个离散随机变量 X 和 Y , 互信息则为:

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (2)$$

$I(X, Y)$ 的值总是非负的, 当且仅当 X 和 Y 独立时 $I(X, Y) = 0$ 。

设 $\{(x_i, y_i), 1 < i < n\}$ 是来自样本量为 n 的随机向量 (X, Y) 的样本。假设 (X, Y) 具有连续的概率密度函数, 记 $p(x), p(y)$ 分别为 x, y 的概率密度函数, $p(x, y)$ 是 x 和 y 的联合概率密度函数。由于对于两个连续随机变量 X 和 Y 的互信息计算时既含有积分是运算, 又含有未知的密度函数, 因此, 直接对其估计是十分困难的, 再由随机变量的数学期望公式, 对于两个连续随机变量 X 和 Y 的互信息可以写成如下形式。

$$I(X, Y) = \iint p(x, y) \log_2 \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy = E \left\{ \log_2 \left(\frac{p(x, y)}{p(x)p(y)} \right) \right\} \quad (3)$$

因此, 当 $E \left\{ \log_2 \left(\frac{p(x, y)}{p(x)p(y)} \right) \right\}$ 存在时, 两个连续随机变量 X 和 Y 的互信息可以考虑采用下式来进行估计:

$$\hat{I}(X, Y) = E \left\{ \log_2 \left(\frac{\hat{p}(x, y)}{\hat{p}(x)\hat{p}(y)} \right) \right\} = \frac{1}{n} \sum_{i=1}^n \log_2 \left(\frac{\hat{p}(x_i, y_i)}{\hat{p}(x_i)\hat{p}(y_i)} \right) \quad (4)$$

其中 $\hat{p}(x)$, $\hat{p}(y)$ 和 $\hat{p}(x, y)$ 分别是 $p(x), p(y)$ 和 $p(x, y)$ 的非参数核密度估计, 具体的有:

$$\hat{p}(x, y) = \frac{1}{nh_1h_2} \sum_{i=1}^n K\left(\frac{x-x_i}{h_1}, \frac{y-y_i}{h_2}\right) = \frac{1}{nh_1h_2} \sum_{i=1}^n K\left(\frac{x-x_i}{h_1}\right) K\left(\frac{y-y_i}{h_2}\right) \quad (5)$$

另外有:

$$\hat{p}(x) = \frac{1}{nh_3} \sum_{i=1}^n K\left(\frac{x-x_i}{h_3}\right) \quad (6)$$

$$\hat{p}(y) = \frac{1}{nh_4} \sum_{i=1}^n K\left(\frac{y-y_i}{h_4}\right) \quad (7)$$

其中, h_1, h_2, h_3 和 h_4 为核密度估计的窗宽, 控制着其光滑程度, 本文根据拇指法则选取 $h_i = 1.06\hat{\sigma}n^{-1/5}, i=1, 2, 3, 4$, 其中 $\hat{\sigma}$ 为数据集的样本标准差。 $K(\cdot)$ 为核函数, 本文取高斯核函数为:

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \quad (8)$$

2.2. 基于互信息的方法

设有 n 个 p 维的特征向量来自 k 个类别, 矩阵表示为:

$$X_i = [x_1^i, x_2^i, \dots, x_p^i] \quad i = 1, 2, \dots, n \quad (9)$$

两个变量 x_i 和 x_j 的互信息量定义为:

$$I(x_i, x_j) = \iint p(x_i, x_j) \log_2 \left(\frac{p(x_i, x_j)}{p(x_i)p(x_j)} \right) dx_i dx_j \quad (10)$$

该互信息值可由上述介绍方法进行估计, 即

$$I_{x_i, x_j} = E \left\{ \log_2 \left(\frac{p(x_i, x_j)}{p(x_i)p(x_j)} \right) \right\} \quad (11)$$

然后估计所有特征对的所有互信息值, 并计算一个特征在特征上的平均绝对互信息:

$$I_{j, \delta} = \frac{1}{\delta} \sum_{i=1, i \neq j}^{\delta} |I_{x_i, x_j}| \quad (12)$$

计算平均互信息值最大的特征:

$$\alpha = \arg \max_j I_{j, \delta} \quad (13)$$

在特征选择算法的每个迭代步骤中每次的平均互信息值最大的特征将被删除。当特征从特征集中移除时, 它也从剩余的平均相关性中被丢弃, 即:

$$I_{j, \delta-1} = \frac{\delta I_{j, \delta} - |I_{x_\alpha, x_j}|}{\delta - 1} \quad (14)$$

2.3. 无监督的特征选择方法

根据上述描述, 建立基于核密度互信息(Kernel Density Mutual Information, KMI)的无监督特征选择方法, 所提出的方法的基本步骤如下:

- Step 1. 初始化 $\delta = p - 1$ ，计算变量间的互信息值矩阵 I ；
 Step 2. 根据(13)式计算平均最大互信息值；
 Step 3. 丢弃平均互信息值最大的特征；
 Step 4. 更新 $\delta = \delta - 1$ ，如果 $\delta = D$ ，则返回生成的个特征子集 D 并停止，否则，执行 Step 5；
 Step 5. 根据式(14)计算剩余特征的平均互信息；
 Step 6. 返回 Step 3。
 该算法从原始数据集中产生最优的维特征子集 $X = [x_1, x_2, \dots, x_D]$ 。

3. 实例分析

3.1. 数据集

为了检验所提方法的有效筛选性能，将检验该方法在两个数据集(二分类数据集和多分类数据)上的筛选性能，二分类数据集采用 Golub [9]等人所得到的白血病数据集(ALLAML 数据集)，数据集包含 72 个样本和 7129 个特征，其中响应变量 cancer 包含两类，分别为急性髓系白血病(AML)共 25 例和急性淋巴细胞白血病(ALL)共 47 例，将其转换成数值型变量，即 0-1 变量，其中标签 0 表示未患急性淋巴细胞白血病(ALL)，标签 1 表示患急性髓系白血病(AML)。多分类数据集采用 UCI 的肺癌数据集(lung 数据集)，该数据集包括 203 个样本和 3312 个特征，具有 5 个连续值标签，即为 5 分类的数据集。这两个数据集都可从 <https://jundongli.github.io/scikit-feature/datasets.html> 下载获取，两个数据集的参数汇总如表 1 所示：

Table 1. Summary table of dataset parameters

表 1. 数据集参数汇总表

数据集	样本量	特征数	类别数
ALLAML	72	7129	2
lung	203	3312	5

3.2. 实验设置

将两个数据集分别应用于本文所提出的筛选方法，其中保留的特征数量参考 Fan 和 Lv [10]等人所建议的特征个数，即 $D = 2 * \lceil n / \log(n) \rceil$ ，其中 $\lceil a \rceil$ 表示取 a 的整数部分， n 为样本量，然后将降维后的数据集分别用支持向量机[11] (Support Vector Machine, SVM)、随机森林[12] (Random Forests, RF)和最近邻[13] (K-Nearest Neighbors, KNN)建立分类模型，来评估特征选择方法的性能。在上述方法中，最近邻将 K 设置为 3，支持向量机使用线性核，随机森林采用五折交叉验证且树的数量为 500 棵。

实验分为三部分，第一部分是数据处理部分，为了消除误差，确保计算的有效性，将数据进行了标准化处理，然后将 80% 的数据作为训练集，20% 的数据作为测试集。第二部分为模型建立部分，分别建立 SVM、RF 和 KNN 模型，每个模型分别训练 10 次，特征个数从 2 增加到 D 。第三部分为模型评估部分，在本实验中，我们使用 10 次的平均分类精度和平均 $F1$ 来评估特征子集上的分类效果。分类精度是指正确分类的样本数量占样本总数的比例。其中 $F1$ 的定义如下：

$$F1 = \frac{2 * P * R}{P + R} \quad (15)$$

其中 P 为精确率， R 为召回率，其计算公式分别为：

$$R = \frac{TP}{TP + FN} \quad (16)$$

$$P = \frac{TP}{TP + FP} \tag{17}$$

上述两式中 TP、FP、TN 和 FN 分别为真正例、假正例、真反例和假反例。分类精度并不适用于所有情况，特别是在不平衡的数据集中，在某些情况下，F1 值等可能更加全面地评估模型的性能。F1 可以用来较好的度量二元分类问题，当类别数大于 2 时，可以使用宏平均 F1 将多类分类问题的 F1 值视为个二元分类的平均 F1 值。宏平均 F1 定义为：

$$Macro_F1 = \frac{2 * Macro_P * Macro_R}{Macro_P + Macro_R} \tag{18}$$

其中 Macro_P 和 Macro_R 分别为宏观精度率和宏观召回率，计算公式分别为：

$$Macro_P = \frac{1}{k} \sum_{i=1}^k P_i \tag{19}$$

$$Macro_R = \frac{1}{k} \sum_{i=1}^k R_i \tag{20}$$

其中 P_i 为第 i 个类别的精度， R_i 为第 i 个类别的召回率， k 表示类别的数量。

3.3. 实验结果与讨论

3.3.1. 分类精度

为了证明本文提出的 KMI 算法的筛选性能，本节将文中提出的 KMI 算法与 Haindl [8]等人提出基于互相关性的方法进行了比较。分别使用 RF、SVM 和 KNN 分类器对两种特征选择算法筛选的特征进行训练，并利用分类精度来衡量不同特征选择算法的分类性能。

表 2 记录了 KMI 算法与基于互相关性算法两种特征选择算法在 ALLAML 数据集和 lung 数据集上的平均分类精度。在每个方法中用粗体标记出最大的平均分类精度。由表 2 可知，在 ALLAML 数据集中，采用 KMI 方法进行变量筛选后，RF 分类模型的平均分类精度达到了 76.0%，而在其余两中方法上，由 KMI 方法筛选后比用互相关性方法筛选后建立的分类模型效果略低，且在三种分类模型中，SVM 分类器的效果最好，分别达到了 79.3%和 78.0%。在 lung 数据集中，分类模型 RF 和 KNN 在使用 KMI 方法筛选后的分类效果都比使用互相关性的方法筛选后的效果要好，因此，与基于互相关性方法相比，KMI 算法更能提高分类模型的性能。

Table 2. Classification accuracy of three classification algorithms on two data sets

表 2. 三种分类算法在两个数据集上的分类精度

	ALLAML		lung	
	Correlation	KMI	Correlation	KMI
RF	0.720	0.760	0.893	0.893
SVM	0.793	0.780	0.951	0.941
KNN	0.730	0.717	0.900	0.924

为了观察特征子集大小对分类精度的影响，比较两种不同特征选择算法在不同保留特征数下的分类精度，绘制了在两个数据集上三个分类方法保留特征数的分类精度图，其横轴表示特征子集中保留的特征数量，纵轴表示分类精度，在图中，使用不同的颜色来区分不同的特征选择方法。

图 1 绘制了三种分类方法在 ALLAML 数据集上的平均分类准确率图，图 2 显示了三种分类方法在 lung 数据集上的平均分类准确率图，由两图可知，模型的分类精度随保留变量的数量增大而逐渐提升，在 lung 数据集中，当保留变量个数在 25 个前，模型的分类精度增加较快，当保留变量个数达到 25 个后模型的分类效果呈缓慢的状态增加，且当保留变量数为 25 个时在三个分类模型中的分类准确率已经接近 90%，且从整体上看，基于 KMI 方法的分类效果要优于基于互相关性方法的效果。而在 ALLAML 数据集中，当保留变量数高于 15 个时，基于 KMI 方法的分类效果总体略高于基于互相关性的方法。

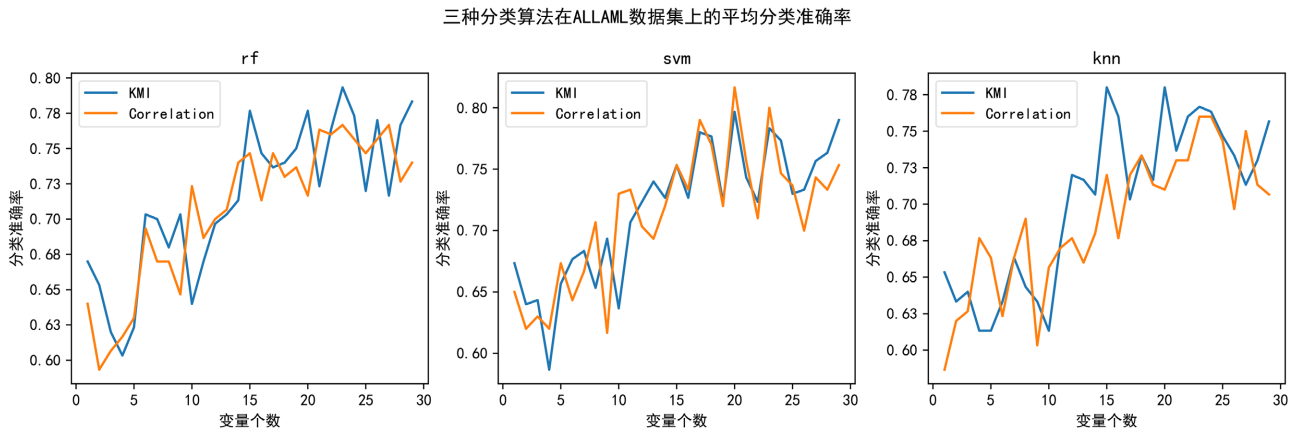


Figure 1. Average classification accuracy of three classification methods on the ALLAML dataset

图 1. 三种分类方法在 ALLAML 数据集上的平均分类准确率

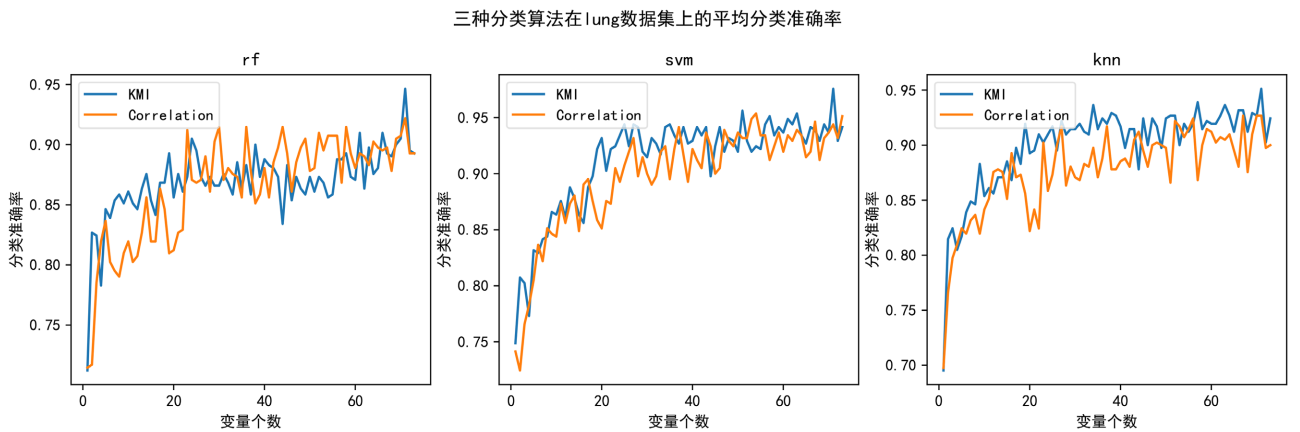


Figure 2. Average classification accuracy of three classification methods on the lung dataset

图 2. 三种分类方法在 lung 数据集上的平均分类准确率

3.3.2. F1 值

F1 值是精确率和召回率的调和平均值，适用于类不平衡问题的性能评价。为了评价本文提出的 KMI 算法的筛选性能，我们使用 F1 值来评价在二分类数据集上的分类器的结果，使用宏 F1 值来评价在多分类数据集上的分类器的结果，表 3 记录了 KMI 算法与基于互相关性的方法两种特征选择算法在 ALLAML 数据集上的平均 F1 值和 lung 数据集上的宏平均 F1 值。由表三数据可知，在 ALLAML 数据集上，RF 和 SVM 两个分类模型在 KMI 算法上的性能更好，在 KNN 上略低于基于互相关性的方法，在 lung 数据集上，三个模型在基于 KMI 上的 F1 值都比基于互相关性的方法要好。更进一步说明了 KMI 算法更能提高分类模型的性能。

Table 3. F1 values of the three classification algorithms in the two datasets
表 3. 三种分类算法在两个数据集 F1 值

	ALLAML		lung	
	Correlation	KMI	Correlation	KMI
RF	0.802	0.837	0.750	0.805
SVM	0.835	0.837	0.864	0.899
KNN	0.811	0.803	0.789	0.861

同时也绘制了在不同保留特征数量下三种模型的平均 F1 值和宏平均 F1 值，以比较两种不同特征选择算法在保留特征数条件下 F1 值的变化，如图 3 和图 4 所示，图 3 显示了三种分类方法在 ALLAML 数据集上的平均 F1 值，图 4 绘制了三种分类方法在 lung 数据集上的宏平均 F1 值。由图可知，其变化规律同三个分类模型的分类精度变化趋势类似。

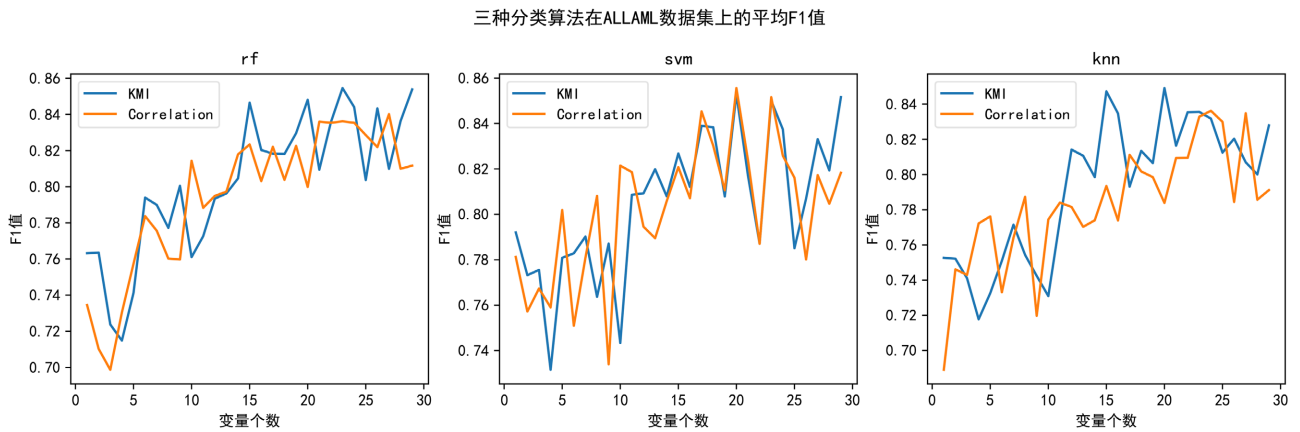


Figure 3. Average F1 value of three classification methods on the ALLAML dataset
图 3. 三种分类方法在 ALLAML 数据集上的平均 F1 值

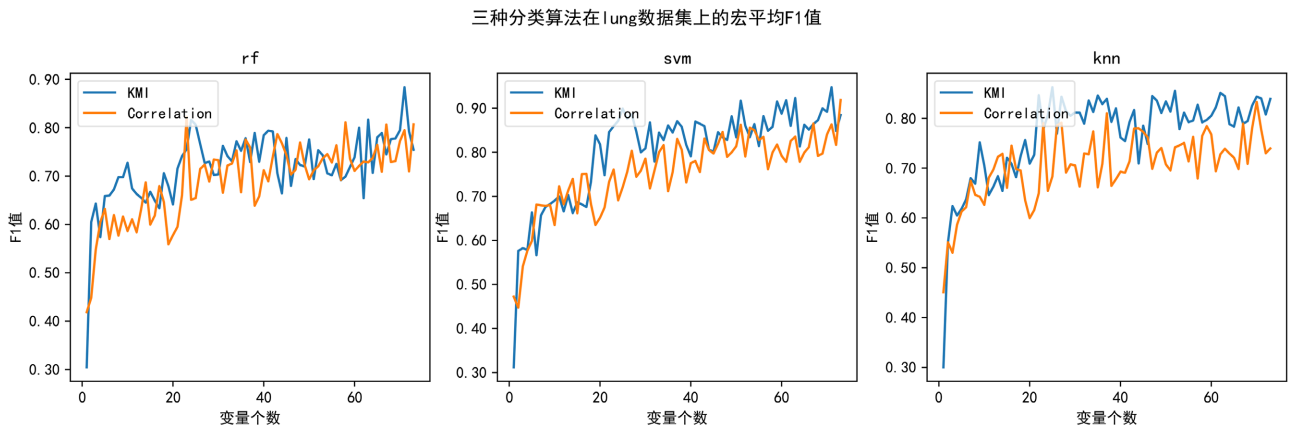


Figure 4. Macro_F1 values of the three classification methods on the lung dataset
图 4. 三种分类方法在 lung 数据集上的宏平均 F1 值

4. 总结

本文主要研究了基于核密度互信息的无监督特征选择算法。研究的目的是去除特征之间冗余度较高

的特征子集，以实现降维。作为一种过滤方法，它产生独立于要使用的分类器的特征。同时该方法引入互信息作为特征之间的相关性度量，可以很好的度量特征间的线性和非线性关系，采用核密度估计的方式估计特征的密度函数，可以对互信息更为准确的计算。文章通过两个高维数据集验证了所提方法的筛选性能，并将其与现有的基于互相关性方法作对比，最后将筛选后的数据应用于三个常用的分类模型，用分类精度和平均 F1 值对分类模型进行评价，实验结果表明，在两个数据集上，两种筛选方法在三个分类模型中的分类精度表现类似，而基于核密度互信息的方法筛选后应用三种分类算法的平均 F1 值比基于互相关性的方法表现更好，表明了本文所提方法具有更好的筛选性能。

参考文献

- [1] Zhou, H., Wang, X. and Zhu, R. (2022) Feature Selection Based on Mutual Information with Correlation Coefficient. *Applied Intelligence*, **52**, 5457-5474. <https://doi.org/10.1007/s10489-021-02524-x>
- [2] Solorio-Fernández, S., Carrasco-Ochoa, J.A., and Martínez-Trinidad, J.F. (2020) A Review of Unsupervised Feature Selection Methods. *Artificial Intelligence Review*, **53**, 907-948. <https://doi.org/10.1007/s10462-019-09682-y>
- [3] Hancer, E., Xue, B. and Zhang, M. (2020) A Survey on Feature Selection Approaches for Clustering. *Artificial Intelligence Review*, **53**, 4519-4545. <https://doi.org/10.1007/s10462-019-09800-w>
- [4] Hancer, E. (2020) A New Multi-Objective Differential Evolution Approach for Simultaneous Clustering and Feature Selection. *Engineering Applications of Artificial Intelligence*, **87**, Article ID: 103307. <https://doi.org/10.1016/j.engappai.2019.103307>
- [5] Zhu, Q.H. and Yang, Y.B. (2018) Discriminative Embedded Unsupervised Feature Selection. *Pattern Recognition Letters*, **112**, 219-225. <https://doi.org/10.1016/j.patrec.2018.07.018>
- [6] Solorio-Fernández, S., Martínez-Trinidad, J.F. and Carrasco-Ochoa, J.A. (2017) A New Unsupervised Spectral Feature Selection Method for Mixed Data: A Filter Approach. *Pattern Recognition*, **72**, 314-326. <https://doi.org/10.1016/j.patcog.2017.07.020>
- [7] Haindl, M., Somol, P., Verwerdis, D., et al. (2006) Feature Selection Based on Mutual Correlation. In: Martínez-Trinidad, J.F., Carrasco Ochoa, J.A. and Kittler, J., Eds., *Progress in Pattern Recognition, Image Analysis and Applications, CIARP 2006*, Springer, Berlin, Heidelberg, 569-577. https://doi.org/10.1007/11892755_59
- [8] Zhou, S., Wang, T. and Huang, Y. (2022) Feature Screening via Mutual Information Learning Based on Nonparametric Density Estimation. *Journal of Mathematics*, **2022**, Article ID: 7584374. <https://doi.org/10.1155/2022/7584374>
- [9] Golub, T.R., Slonim, D.K., Tamayo, P., et al. (1999) Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science*, **286**, 531-537. <https://doi.org/10.1126/science.286.5439.531>
- [10] Fan, J. and Lv, J. (2008) Sure Independence Screening for Ultrahigh Dimensional Feature Space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70**, 849-911. <https://doi.org/10.1111/j.1467-9868.2008.00674.x>
- [11] Land, W.H., Schaffer, J.D., Land, W.H., et al. (2020) The Support Vector Machine. In: Land, W.H. and Schaffer, J.D., Eds., *The Art and Science of Machine Intelligence: With an Innovative Application for Alzheimer's Detection from Speech*, Springer, Cham, 45-76. https://doi.org/10.1007/978-3-030-18496-4_2
- [12] Breiman, L. (2021) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>
- [13] Peterson, L.E. (2009) K-Nearest Neighbor. *Scholarpedia*, **4**, Article No. 1883. <https://doi.org/10.4249/scholarpedia.1883>